

StableRep: Synthetic Images from Text-to-Image Models Make Strong Visual Representation Learners

Year: 2023 | Citations: 203 | Authors: Yonglong Tian, Lijie Fan, Phillip Isola, Huiwen Chang, Dilip Krishnan

Abstract

We investigate the potential of learning visual representations using synthetic images generated by text-to-image models. This is a natural question in the light of the excellent performance of such models in generating high-quality images. We consider specifically the Stable Diffusion, one of the leading open source text-to-image models. We show that (1) when the generative model is configured with proper classifier-free guidance scale, training self-supervised methods on synthetic images can match or beat the real image counterpart; (2) by treating the multiple images generated from the same text prompt as positives for each other, we develop a multi-positive contrastive learning method, which we call StableRep. With solely synthetic images, the representations learned by StableRep surpass the performance of representations learned by SimCLR and CLIP using the same set of text prompts and corresponding real images, on large scale datasets. When we further add language supervision, StableRep trained with 20M synthetic images achieves better accuracy than CLIP trained with 50M real images.